MA8020 Tekniska beräkningar

Något om optimering

Mikael Hindgren

HÖGSKOLAN I HALMSTAD

4 december 2025

Optimering
Grundprinciper

- *Definition:* Användning av matematiska modeller och metoder för att finna det ”bästa alternativet” - *Optimum*
- *Målet:* Kunskaper som krävs för att fatta det bästa möjliga beslutet
- *Standardmodellens komponenter:*
  - *Målfunktion*: f ska minimeras/maximeras
  - *Designvariabler*: Variablerna som styr resultatet
- *Motivering:*
  - Minimera vikt
  - Maximera resursallokering
  - Minimera energiförbrukning
  - Maximera vinst
  - Minimera tid
  - Maximera prestanda
  - Minimera risk
  - ...

Optimering

Begränsningar och modellers komplexitet

- *Obegränsad optimering:* Designvariablerna är fria (sällan i praktiken).
- *Begränsad optimering* (vanligast):
  - Inkluderar naturliga begränsningar sk bivillkor ($g(x) \leq 0$).
  - Lösningen söks i den *tillåtna regionen* (feasible region).
  - Form: Minimera/Maximera $f(x)$ under bivillkor/begränsningar (s.t.).
- *Modellernas komplexitet:*
  - *Linjär Programmering (LP):* Linjär målfunktion och linjära begränsningar.
  - *Icke-linjära modeller (NP):* Målfunktion och/eller begränsningar är icke-linjära $\Rightarrow$ svårare att hantera.
- *Utmaning:* Inga generella garantier kan ges för att uppnå ett *globalt optimum* (eller ens ett lokalt) i de mest allmänna modellerna.

Optimering

Optimum hos kontinuerliga funktioner

Sats 1 (Satsen om största och minsta värde)

Om $D \subset \mathbb{R}^n$ är en kompakt mängd och $f : D \to \mathbb{R}$ en kontinuerlig funktion på D så gäller följande:

- *f antar sitt största $(\max f)$ och minsta $(\min f)$ värde i D (Extremvärdessatsen).*
- *Den punkt $\mathbf{x}_0 \in D$ där ett extremvärde antas är av följande typ*
  1. *Kritisk punkt i det inre av D där $\nabla f(\mathbf{x}_0) = \mathbf{0}$.*
  2. *Singulär punkt i det inre av D där $\nabla f(\mathbf{x}_0)$ är odefinierad.*
  3. *Randpunkt: $\mathbf{x}_0 \in \partial D$.*

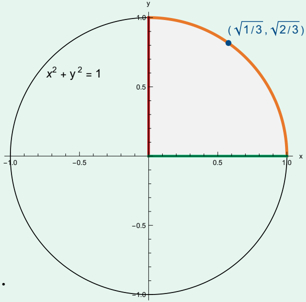
Optimering

Exempel 1

Bestäm maximum för $f(x,y) = xy^2$ då $x^2 + y^2 \leq 1$, $x \geq 0$, $y \geq 0$.

- Kritiska punkter: $\nabla f = (y^2, 2xy) = (0,0) \Rightarrow (x,y) = (0,0)$
- Randpunkter:

$$\begin{aligned}
x = 0 \quad &\Rightarrow \quad f(x,y) = g_1(y) = 0 \\
y = 0 \quad &\Rightarrow \quad f(x,y) = g_2(x) = 0 \\
y^2 = 1 - x^2 \quad &\Rightarrow \quad f(x,y) = g_3(x) = x(1-x^2) \\
&\Rightarrow \quad g_3'(x) = -3x^2 + 1 = 0 \\
&\Leftrightarrow \quad (x,y) = \left(\pm\sqrt{\frac{1}{3}}, \sqrt{\frac{2}{3}}\right) \\
&\therefore \quad f_{\max} = f\left(\sqrt{\frac{1}{3}}, \sqrt{\frac{2}{3}}\right) = \frac{2}{3\sqrt{3}}.
\end{aligned}$$



Optimering
Riktingsderivata

Definition 1

Om $\boldsymbol{v}$ är en enhetsvektor ($|\boldsymbol{v}| = 1$) är riktningsderivatan av funktionen $f(\boldsymbol{x})$ i punkten $\boldsymbol{a}$ i riktning $\boldsymbol{v}$

$$\frac{df}{d\boldsymbol{v}}(\boldsymbol{a}) = \lim_{h\to 0} \frac{f(\boldsymbol{a}+h\boldsymbol{v}) - f(\boldsymbol{a})}{h}$$

Kedjeregeln ger ett mer användbart uttryck:

$$\frac{df}{d\boldsymbol{v}}(\boldsymbol{a}) = \nabla f(\boldsymbol{a}) \cdot \boldsymbol{v}$$

Exempel 2

Bestäm riktingsderivatan av $f(x, y) = x^2 + 2xy - 3y^2$ i punkten $\boldsymbol{a} = (1, 2)$ i riktning $\boldsymbol{v} = \frac{1}{\sqrt{2}}(1, -1)$.

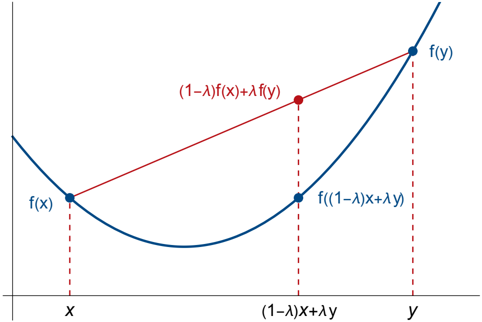
$$\begin{aligned} \nabla f &= \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) = (2x + 2y, 2x - 6y) \\ \Rightarrow \nabla f(1, 2) \cdot \boldsymbol{v} &= (6, -10) \cdot \frac{1}{\sqrt{2}}(1, -1) = 8\sqrt{2}. \end{aligned}$$

Definition 2 (Konvex, konkav)

- Funktionen $f : C \to \mathbb{R}$ är konvex om $x, y \in C$, $\lambda \in [0, 1]$ om

$$f((1-\lambda)x, \lambda y) \leq (1-\lambda)f(x) + \lambda f(y)$$

- Om olikheten är strikt ($<$) är f strikt konvex.
- Om $-f$ är (strikt) konvex är f (strikt) konkav.

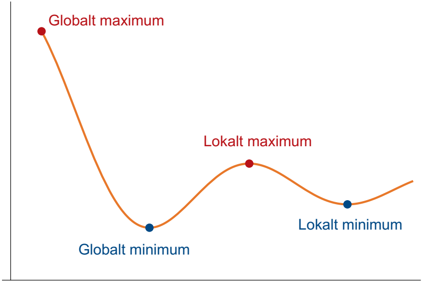


Definition 3

Funktion f har ett:

- globalt maximum i x^* om $f(x^*) \geq f(x)$ för alla $x \in D_f$,
- lokalt maximum om det finns liten omgivning U av x^* så att $f(x^*) \geq f(x)$ för $x \in U$.

Om olikheten är sträng ($<$) är det ett strikt maximimum.
Globalt och lokalt minimum definieras på motsvarande sätt.



Optimering

Maxima och minima

När är en stationär punkt $\boldsymbol{x}_0$ en lokal maximi, minimi- eller sadelpunkt?
Taylorutveckling i flera variabler:

$$f(\boldsymbol{x}) = f(\boldsymbol{x}_0) + \underbrace{(\nabla f(\boldsymbol{x}_0))}_{=0}^{T}(\boldsymbol{x} - \boldsymbol{x}_0) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_0)^T H(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) + \ldots$$

$$H(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix} \leftarrow \text{Hessianska matrisen ("hessianen")}$$

$$\Rightarrow f(\boldsymbol{x}) - f(\boldsymbol{x}_0) \approx \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_0)^T H(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0)$$

Tecknet på $f(\boldsymbol{x}) - f(\boldsymbol{x}_0)$ dvs. om $f(\boldsymbol{x}) \geq f(\boldsymbol{x}_0)$ eller $f(\boldsymbol{x}) \leq f(\boldsymbol{x}_0)$ beror på H.

Optimering

Maxima och minima

Definition 4

Om A är en $n \times n$ -matris och $\boldsymbol{v} \in \mathbb{R}^n$ så är A

- positivt definit om $\boldsymbol{v}^T A \boldsymbol{v} > 0$ for alla $\boldsymbol{v} \neq 0$,
- positivt semidefinit om $\boldsymbol{v}^T A \boldsymbol{v} \geq 0$ for alla $\boldsymbol{v} \neq 0$,
- negativt definit om $\boldsymbol{v}^T A \boldsymbol{v} < 0$ for alla $\boldsymbol{v} \neq 0$,
- negativt semidefinit om $\boldsymbol{v}^T A \boldsymbol{v} \leq 0$ for alla $\boldsymbol{v} \neq 0$,
- indefinit om A varken är positivt eller negativt semidefinit.

Sats 2

- *Om $\nabla f(\boldsymbol{x}_0) = 0$ och $H(\boldsymbol{x}_0)$ är positivt (definit) semidefinit så har f ett (strikt) lokalt minimum i $\boldsymbol{x}_0$.*
- *Om f är konvex och $\nabla f(\boldsymbol{x}_0) = 0$ har f globalt minimum i $\boldsymbol{x}_0$.*

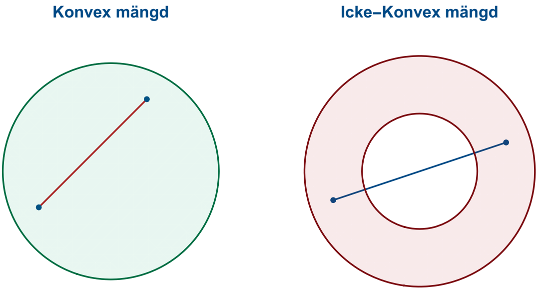
Optimering

Konvexa mängder

Definition 5

En mängd $D \subset \mathbb{R}^n$ är konvex om, för varje par av punkter $\boldsymbol{x}$ och $\boldsymbol{y}$ i mängden, hela det räta linjesegmentet mellan $\boldsymbol{x}$ och $\boldsymbol{y}$ också ligger inuti D:

$$D \text{ är konvex } \Leftrightarrow \forall\, \boldsymbol{x}, \boldsymbol{y} \in D, \forall \lambda \in [0,1] : \; \lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y} \in D$$



Optimering

Konvexa funktioner

Sats 3

Låt M vara en konvex mängd, f deriverbar två gånger och H f:s hessian.

- *$f(x)$ är en konvex funktion på M om H är positivt semidefinit*
- *$f(x)$ är en strikt konvex funktion på M om H är positivt definit*
- *$f(x)$ är en konkav funktion på M om H är negativt semidefinit*
- *$f(x)$ är en strikt konkav funktion på M om H negativt definit*

Sats 4

Låt λ_i vara egenvärdena till hessianen H. Om alla

- *$\lambda_i \geq 0$ är H positivt semidefinit*
- *$\lambda_i > 0$ är H positivt definit*
- *$\lambda_i \leq 0$ är H negativt semidefinit*
- *$\lambda_i < 0$ är H negativt definit*

Optimering

Problemformulering och villkor för lokala och globala minima.

Ett allmänt optimeringsproblem kan formuleras som

$$(G) \qquad \begin{array}{ll} \min & f(x) \\ \text{s.t.} & x \in F \quad \text{(s.t. "subject to")} \end{array}$$

där $f : F \to \mathbb{R}$ och $F \subseteq \mathbb{R}^n$.

- F kallas lösningsområde (feasible region) och f målfunktion (objective function)
- Om $F \subset \mathbb{R}^n$ har vi optimering med bivillkor och om $F = \mathbb{R}^n$ optimering utan bivillkor.
- I explicit matematisk form kan (G) representeras som

$$(M) \qquad \begin{array}{ll} \min & f(x) \\ \text{s.t.} & \begin{cases} g_i(x) \leq 0, \; i = 1, ..., m \\ h_j(x) = 0, \; j = 1, ..., p \\ x \in \mathbb{R}^n \end{cases} \end{array}$$

dvs $F = \{x \in \mathbb{R}^n \mid g_i(x) \leq 0, \; h_j(x) = 0\}$.

Optimering

Kuhn-Tucker-villkor för lokalt minimum.

Sats 5 (Kuhn-Tucker-villkoren)

För det matematiska optimeringsproblemet:

$$(M) \qquad \begin{array}{ll} \min & f(x) \\ \text{s.t.} & \begin{cases} g_i(x) \leq 0, \ i = 1, \ldots, m \\ h_j(x) = 0, \ j = 1, \ldots, p \\ x \in \mathbb{R}^n \end{cases} \end{array}$$

är följande villkor nödvändiga för lokalt minimum i x^* om ett lämpligt regularitetsvillkor (t.ex. LICQ) är uppfyllt (KT 1-4):

1. *Primal tillåtlighet:* $g_i(x^*) \leq 0, \ h_j(x^*) = 0$
2. *Komplementaritet:* $\lambda_i g_i(x^*) = 0$
3. *Dual tillåtlighet:* $\lambda_i^* \geq 0, \ \mu_j^* \in \mathbb{R}$
4. *Stationaritet:* $\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^{p} \mu_j^* \nabla h_j(x^*) = \mathbf{0}$

Optimering

Kuhn-Tucker-villkor för lokalt minimum.

Förklaringar:

- Primal tillåtlighet: Lösningen x^* måste vara tillåten (feasible) i det ursprungliga primala problemet (M).
- Komplementaritet: Villkoret $\lambda_i^* g_i(x^*) = 0$ innebär att antingen är multiplikatorn λ_i^* noll, eller så är olikhetsbivillkoret $g_i(x^*)$ aktivt (mättat).
- Dual tillåtlighet: Lagrangemultiplikatorerna λ_i^* för olikhetsbivillkoren måste vara icke-negativa ($\lambda_i^* \geq 0$).
- Stationaritet: Villkoret $\nabla_{\mathbf{x}} L(x^*, \lambda^*, \mu^*) = \mathbf{0}$ säger att x^* är en stationär punkt till Lagrangefunktionen L.

Anm:

- Om endast likhetsvillkor ($h_j(x) = 0$) finns, reduceras KKT-villkoren till *Lagrangevillkoren*: $h_j(x^*) = 0$ och Stationaritetsvillkoret (4).
- Om M är ett *konvext optimeringsproblem* (konvex f, konvexa g_i, linjära h_j), är KT-villkoren *tillräckliga* för att garantera ett globalt minimum i x^*, under förutsättning att $\lambda^* \geq 0$ (Dual tillåtlighet) är uppfyllt.

Optimering utan bivillkor

Polynomapproximation

Förutsättningar:

- $f(x)$ är konvex i intervallet $I = [x_1, x_3] \Rightarrow$ har minimum
- Funktionsvärden kan beräknas för alla $x \in I$ och $x_1 < x_2 < x_3$

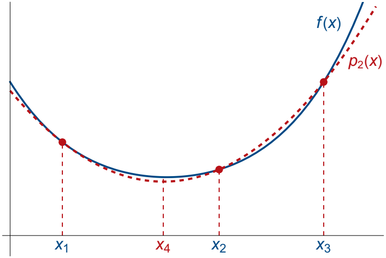
Approximera $f(x)$ med ett polynom $p_2(x)$ av grad 2:

$$\begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} \Leftrightarrow p_2(x) = c_0 + c_1 x + c_2 x^2 \Rightarrow p_2'(x) = c_1 + 2c_2 x$$

$$\Rightarrow x_4 = -\frac{c_1}{2c_2} = \dots = \frac{f_1(x_2^2 - x_3^2) + f_2(x_3^2 - x_1^2) + f_3(x_1^2 - x_2^2)}{2(f_1(x_2 - x_3) + f_2(x_3 - x_1) + f_3(x_1 - x_2))}$$

Iteration:

1. Om $x_4 < x_2$:
   - Om $f(x_4) < f(x_2)$: $x_3 = x_2,\ x_2 = x_4$
   - Annars $(f(x_4) \geq f(x_2))$: $x_1 = x_4$
2. Annars $(x_4 > x_2)$:
   - Om $f(x_4) < f(x_2)$: $x_1 = x_2,\ x_2 = x_4$
   - Annars $(f(x_4) \geq f(x_2))$: $x_3 = x_4$
3. Fortsätt tills $|x_3 - x_1| < \varepsilon$



Optimering utan bivillkor

Polynomapproximation

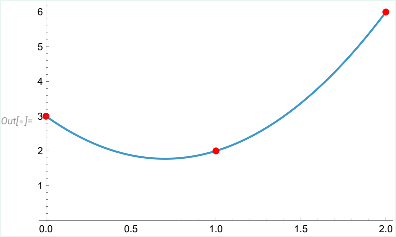
Exempel 3

Följande data är given:

x	0	1	2
f	3	2	6

Bestäm ett approximativt värde på minimum för $f(x)$ i $[0, 2]$ med hjälp av $p_2(x)$.

Mathematica: $p_2(x) = 3. - 3.5x + 2.5x^2$, $x_4 = 0.7$, $f_{min} \approx 1.775$



Optimering utan bivillkor

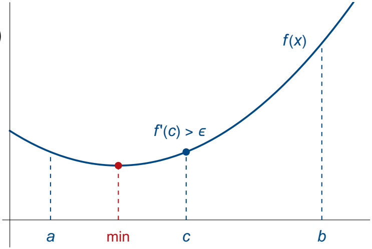
Bisektionsmetoden (Bolzano Search)

Förutsättningar:

- $f(x)$ är konvex i $[a, b]$
- $f'(x)$ är kontinuerlig i $[a, b]$ (och kan beräknas)
- $f'(a)f'(b) < 0$

Iteration:

1. Beräkna $f'(c)$ där $c = \frac{b+a}{2}$
2. Om $|f'(c)| < \varepsilon$ accepteras c som minimum
   Annars: Om $f'(c) > \varepsilon$, $b = c$ annars $a = c$
3. Iterera tills $|f'(c)| < \varepsilon$



Metoden konvergerar till globalt minimum om f är konvex (i hela D_f), annars till ett lokalt minimum om det är en stationär punkt.

Exempel 4

Minimera $f(x) = x^2 - \dfrac{4x}{3} + 1$ med Bisektionsmetoden.

Optimering utan bivillkor

Newtons metod

På samma sätt som vi kan hitta en rot till $f(x) = 0$ med Newton-Raphsons metod kan vi hitta en rot till $f'(x) = 0$. Tillräckliga villkor för konvergens:

- $f(x) \in C^2$ i $[a, b]$
- Ett lokalt minimum x^* måste existera tillräckligt nära startpunkten x_0
- $f(x)$, $f'(x)$ och $f''(x)$ kan beräknas i $[a, b]$.
- $f''(x) > 0$ i en omgivning av x^*

Iterationsformel:

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

Exempel 5

Minimera $f(x) = x^2 - \dfrac{4x}{3} + 1$ med Newtons metod.

Anm: Sekantmetoden är ett alternativ som undviker beräkning av $f''(x)$:

$$x_{k+1} = x_k - f'(x_k)\frac{x_k - x_{k-1}}{f'(x_k) - f'(x_{k-1})}$$

Optimering utan bivillkor: Funktioner av flera variabler

Sökmetoder

En vanlig grupp av metoder for optimering är sökmetoder. De kan alla formuleras enligt följande iteration

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{d}^{(k)}, k = 0, 1, 2, 3...$$

där

- $x^{(0)}$ är startpunkt
- α_k är steglängd
- $\boldsymbol{d}^{(k)}$ är sökriktning

Valet av α_k och $\boldsymbol{d}^{(k)}$ definierar de olika metoderna.

Om f är en kvadratisk funktion, dvs. kan skrivas som

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T H \boldsymbol{x} + \boldsymbol{c}^T \boldsymbol{x} + b \quad \text{är} \quad \alpha_k = -\frac{\nabla f(\mathbf{x}_k)^T \mathbf{d}_k}{\mathbf{d}_k^T H \mathbf{d}_k} \quad (H \text{ konstant})$$

- Det är det enklaste icke-linjära fallet där linjesökningsproblemet (att hitta optimalt α_k) har en exakt analytisk lösning.
- Alla C^2-funktioner kan approximeras med en kvadratisk funktion (Taylor-utveckling) och fallet representerar därför det lokala beteendet hos alla optimeringsproblem.

Optimering utan bivillkor: Funktioner av flera variabler

Sökmetoder

- *Steglängden:* I det allmänna fallet måste α_k bestämmas approximativt.
- *Linjesökningsproblemet:* När sökriktningen $\mathbf{d}_k$ har valts, reduceras problemet till ett endimensionellt problem: Hitta det α som minimerar funktionen längs den valda riktningen:

$$\min_{\alpha} f(\mathbf{x}_k + \alpha \mathbf{d}_k) \tag{1}$$

- Eftersom (1) är ett optimeringsproblem i en variabel (α) utan bivillkor, kan tidigare 1D-metoder användas t.ex. kvadratisk interpolation.

Optimering utan bivillkor: Funktioner av flera variabler

Brantaste lutningsmetoden (Steepest Descent)

Metoden använder $\alpha_k = 1$ och $\boldsymbol{d}^{(k)} = -\nabla f(\boldsymbol{x}^{(k)})$:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{d}^{(k)} = \boldsymbol{x}^{(k)} - \nabla f(\boldsymbol{x}^{(k)}), \quad k = 0, 1, 2, 3, ...$$

- Gradienten $\nabla f(\boldsymbol{x}_k)$ pekar i den riktning där funktionen f ökar snabbast i punkten $\boldsymbol{x}_k$.
- Vektorns norm (längd), $||\nabla f(\mathbf{x}_k)||$, är lika med den maximala tillväxthastigheten (riktningsderivatan) i denna riktning.

$\Rightarrow -\nabla f(\mathbf{x}_k)$ är riktningen för brantaste lutningen dvs. där funktionen minskar snabbast.

Optimering utan bivillkor: Funktioner av flera variabler

Newtons metod

Metoden är en generalisering av 1D-fallet $\left(x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}\right)$ med

$$\boldsymbol{d}^{(k)} = -H(\boldsymbol{x}^{(k)})^{-1}\nabla f(\boldsymbol{x}^{(k)}) \quad \text{och} \quad \alpha_k = 1$$

vilket ger

$$\boldsymbol{x}^{(k+1)} = \mathbf{x}_k + \alpha_k \mathbf{d}_k = \boldsymbol{x}^{(k)} - H(\boldsymbol{x}^{(k)})^{-1}\nabla f(\boldsymbol{x}^{(k)}), \quad k = 0, 1, 2, 3, ...$$

- Newton's metod är en direkt generalisering av Newton-Raphsons metod för rotsökning och löser systemet av icke-linjära ekvationer: $\nabla f(\mathbf{x}) = \mathbf{0}$.
- Om hessianen H är icke-singulär, erhålls kvadratisk konvergens (mycket snabb).
- Om startpunkten $\boldsymbol{x}_0$ är långt från minimum $\hat{\boldsymbol{x}}$, används ofta en steglängd $0 < \alpha_k < 1$ (bestäms via linjesökning) i de första iterationerna (Dämpad Newtonmetod).
- Samma försiktighet med steglängden ($\alpha_k < 1$) bör även användas för Brantaste lutningsmetoden.

Optimering utan bivillkor: Funktioner av flera variabler

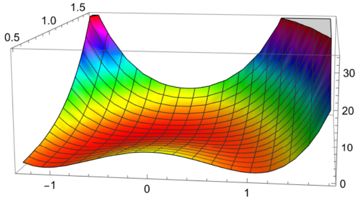
Newtons metod

Exempel 6

Vi minimerar $f(x, y) = (x + 0.3y)^2 + (2(x^2 + y^2 - 1) - 1)^2$

```
In[∘]:= Remove["Global`*"]
        f[x_, y_] := (x + 0.3 y) ^2 + (2 (x ^2 + y ^2 - 1) - 1) ^2;
        Plot3D[f[x, y], {x, -1.3, 1.8}, {y, 0.5, 1.7}, ColorFunction → Hue]
        FindMinimum[f[x, y], {{x, -1.1}, {y, 0.8}}]
```

Out[∘]=

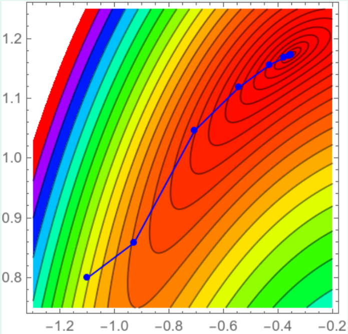


Out[∘]= $\{4.47878 \times 10^{-19}, \{x \to -0.351928, y \to 1.17309\}\}$

Vi väjer $\alpha_k = 0.8$ och startpunkt $\boldsymbol{x}^{(0)} = (-1.1, 0.8)$:

Exempel 6 (Forts.)

```
Remove["Global`*"]
f = (x + 0.3 y) ^2 + (2 (x ^2 + y ^2 - 1) - 1) ^2;
alpha = 0.8;
Print["∇f = ", G = FullSimplify[D[f, {{x, y}}]]];
Print["H = ", H = FullSimplify[D[f, {{x, y}, 2}]]];
{x, y} = {-1.1, 0.8};
route = {{x, y}};
Do[{x, y} = {x, y} - alpha * Inverse[H].G; AppendTo[route, {x, y}], {10}];
Print["Route = ", Transpose[route]];
pl = ContourPlot[f, {x, -1.3, -0.2}, {y, 0.75, 1.25},
   Contours → 6 (Range[0, 6, 0.25] / 6) ^4,
   PlotPoints → 50,
   ColorFunction → Hue,
   Epilog → {Blue, Line[route], PointSize[0.02], Point /@ route}
  ];
Print["fmin = ", f]
Show[pl]
```



$\nabla f = \{16.\,(x^3 + 0.0375\,y + x\,(-1.375 + y^2)),\ 16.\,(0.0375\,x - 1.48875\,y + x^2\,y + y^3)\}$

$H = \{\{-22 + 48\,x^2 + 16\,y^2,\ 0.6 + 32\,x\,y\},\ \{0.6 + 32\,x\,y,\ 48.\,(-0.49625 + 0.333333\,x^2 + y^2)\}\}$

Route = {{-1.1, -0.928934, -0.706371, -0.545316, -0.432761, -0.377824, -0.35827, -0.353271, -0.3522, -0.351982, -0.351939}, {0.8, 0.858773, 1.04653, 1.1184, 1.15598, 1.16842, 1.17203, 1.17287, 1.17305, 1.17308, 1.17309}}

$\text{fmin} = 1.79546 \times 10^{-10}$

Optimering med bivillkor

Lagrangerelaxation

Det matematiska problemet

$$(M) \qquad \begin{aligned} &\min \quad f(x) \\ &\text{s.t.} \begin{cases} g_i(x) \leq 0, \; i = 1, \dots, m \\ x \in X \end{cases} \end{aligned} \tag{2}$$

där X, f och g_i är konvexa kan skrivas om med Lagrangerelaxationen

$$(M_L) \qquad \begin{aligned} &\min \quad \left(f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) \right) \\ &\text{s.t.} \;\; x \in X \end{aligned}$$

- λ_i är Lagrangemultiplikatorer
- Lagrangefunktionen definieras som: $L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$
- Lagrangerelaxationen (M_L) är en relaxation till M om $\lambda_i \geqslant 0$
- Ett nödvändigt villkor för att x^* ska vara en optimal lösning till (2) är att det existerar $\lambda^* \geq 0$ så att stationaritetsvillkoret är uppfyllt:

$$\nabla_x L(x^*, \lambda^*) = \nabla f(x^*) + \sum^{m} \lambda_i^* \nabla g_i(x^*) = \mathbf{0}$$

Optimering med bivillkor

Lagranges metod för likhetsvillkor

Vi ska nu studera problemet

$$(L) \qquad \begin{aligned} &\min \quad f(x) \\ &\text{s.t.} \begin{cases} g_i(x) = 0, \; i = 1, \ldots, m \\ x \in X \end{cases} \end{aligned} \tag{3}$$

och tillhörande Lagrangefunktion:

$$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$$

- För likhetsvillkor (3) finns ingen teckenrestriktion på λ_i ($\lambda_i \in \mathbb{R}$).
- För att hitta stationära punkter behöver vi lösa ekvationssystemet

$$\nabla_{x,\lambda} L(x, \lambda) = \mathbf{0} \Leftrightarrow \begin{cases} \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) = \mathbf{0}, \\ g_i(x) = 0, \; i = 1, \ldots, m. \end{cases}$$

Optimering med bivillkor

Lagranges metod för likhetsvillkor

Exempel 7

Bestäm den punkt på planet $x - y + 2z = 6$ som är närmast origo.
Avståndet från punkten (x, y, z) till origo är $d = \sqrt{x^2 + y^2 + z^2}$.
Optimeringsproblem:

$$(L) \quad \begin{array}{ll} \min & f(x,y,z) = x^2 + y^2 + z^2 \\ \text{s.t.} & g(x,y,z) = x - y + 2z - 6 = 0 \end{array}$$

$$\Rightarrow \quad L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \lambda g(\boldsymbol{x}) = x^2 + y^2 + z^2 + \lambda(x - y + 2z - 6)$$

Vi får

$$\nabla_{\boldsymbol{x},\lambda} L(\boldsymbol{x}, \lambda) = \mathbf{0} \quad \Leftrightarrow \quad \begin{cases} \nabla f(\boldsymbol{x}) + \lambda \nabla g(\boldsymbol{x}) = \mathbf{0} \\ g(\boldsymbol{x}) = 0 \end{cases}$$

$$\Leftrightarrow \quad \begin{cases} (2x + \lambda, 2y - \lambda, 2z + 2\lambda) = \mathbf{0} \\ x - y + 2z - 6 = 0 \end{cases}$$

$$\Leftrightarrow (x, y, z, \lambda) \quad = \quad (1, -1, 2, -2) \Rightarrow f_{\min} = f(1, -1, 2) = 6 \Rightarrow d = \sqrt{6}$$

Optimering med bivillkor

Lagranges metod för likhetsvillkor

Exempel 8

Lös optimeringsproblemet med Newtons metod:

$$(L)\quad \begin{array}{ll} \text{min} & f(x,y,z) = x^2 + y^2 + z^2 \\ \text{s.t.} & \begin{cases} g_1(x,y,z) = x + y + z + 3 = 0 \\ g_2(x,y,z) = x + 2y - 3z - 6 = 0 \end{cases} \end{array}$$

Vi använder Newtons metod:

$$L(\boldsymbol{v}) = L(x,y,z,\lambda_1,\lambda_2) = f(x,y,z) + \lambda_1 g_1(x,y,z) + \lambda_2 g_2(x,y,z)$$

$$\boldsymbol{v}^{(k+1)} = \boldsymbol{v}^{(k)} - H(\boldsymbol{v}^{(k)})^{-1} \nabla L(\boldsymbol{v}^{(k)}), \quad k = 0, 1, 2, 3, \ldots$$

Resultat från Mahematica: $x^* = (-7, -1, 5)$, $f_{\min} = 75$.
Endast en iteration krävs!

Anm: L är kvadratisk $\Rightarrow$ hessianen H är konstant. Om inte så är fallet måste H^{-1} beräknas i varje steg.

Optimering med bivillkor

Straffmetoder för icke-linjära likhetsvillkor

För $f : \mathbb{R}^n \to \mathbb{R}$ och $g : \mathbb{R}^n \to \mathbb{R}^m$ och optimeringsproblemet

$$(P_E) \quad \begin{array}{ll} \min & f(x) \\ \text{s.t.} & g(x) = 0 \end{array}$$

inför vi en obegränsad straffunktion som bestraffar överträdelse av bivillkoren:

$$P(x, r) = f(x) + \frac{1}{r} g(x)^T g(x)$$

- Straffparametern $r > 0$
- Termen $g(x)^T g(x) = ||g(x)||_2^2$ är ett mått på hur mycket bivillkoren $g(x) = 0$ överträds
- Idén är att minimera $P(x, r)$ utan bivillkor

Optimering med bivillkor

Straffmetoder för icke-linjära likhetsvillkor

- Lösningen hittas genom att lösa en *sekvens av obegränsade problem*:

$$(P_k) \quad \min_{x \in \mathbb{R}^n} P(x, r_k)$$

- Vi väljer en sekvens r_k där $r_k \to 0$ när $k \to \infty$.
- *Konvergens:* Om $x^*(k)$ är lösningen till (P_k), gäller (under lämpliga villkor):

$$x^*(k) \to x^* \quad \text{och} \quad \frac{2}{r_k} g(x^*(k)) \to \lambda^* \quad \text{när } r_k \to 0.$$

Anm:

- Kräver lösning av en *sekvens av problem*.
- När $r_k \to 0$ blir hessianen för $P(\mathbf{x}, r_k)$ extremt illa-konditionerad (stort konditionstal) vilket leder till numerisk instabilitet.

Optimering med bivillkor

Straffmetoder för icke-linjära likhetsvillkor

Exempel 9

Lös följande problem med straffmetod:

$$\begin{array}{ll} \min & f(\boldsymbol{x}) = \frac{1}{2}(x_2^2 - x_1^2) \\ \text{s.t.} & x_1 = 1 \end{array}$$

Vi har $g(\boldsymbol{x}) = x_1 - 1$ och bildar straffunktionen

$$P(\boldsymbol{x}, r) = f(\boldsymbol{x}) + \frac{1}{r} g(\boldsymbol{x})^T g(\boldsymbol{x})) = \frac{1}{2}(x_2^2 - x_1^2) + \frac{1}{r}(x_1 - 1)^2$$

$$\nabla_{\boldsymbol{x}} P = \left(-x_1 + \frac{2(x_1 - 1)}{r}, x_2 \right) = \mathbf{0} \Leftrightarrow \boldsymbol{x}^*(r) = \left(\frac{2}{2 - r}, 0 \right) \to \boldsymbol{x}^* = (1, 0) \text{ då } r \to 0$$

För $r < 2$ är hessianen

$$H(\boldsymbol{x}, r) = P_{\boldsymbol{xx}}(\boldsymbol{x}, r) = \begin{pmatrix} -1 + \frac{2}{r} & 0 \\ 0 & 1 \end{pmatrix}$$

positivt definit $\Rightarrow$ P konvex $\Rightarrow$ $\boldsymbol{x}^*$ globalt minimum och $f_{\min} = -\frac{1}{2}$.

Optimering med bivillkor

Straffmetoder för icke-linjära likhetsvillkor

Anm:

- $\frac{2}{r}g(x^*(r)) = \frac{1}{r}(\frac{2}{2-r} - 1) \to 1 = \lambda^*$ då $r \to 0$
- Konditionstalet: $\kappa(H) = \left|\frac{\mu_{\text{max}}}{\mu_{\text{min}}}\right| = \left|\frac{-1+\frac{2}{r}}{1}\right| \to \infty$ då $r \to 0$ (μ egenvärden)
- Detta är nackdelen med (yttre) straffmetod: När $r \to 0$, konvergerar problemet mot $\boldsymbol{x}^*$ men $\kappa(H) \to \infty$. Detta gör problemet mycket svårt att lösa numeriskt på grund av instabilitet och långsam konvergens.

Exempel 10

Lös optimeringsproblemet med straffmetod:

$$\begin{aligned} \text{min} \quad & f(x,y,z) = x^2 + y^2 + z^2 \\ \text{s.t.} \quad & \begin{cases} g_1(x,y,z) = x + y + z + 3 = 0 \\ g_2(x,y,z) = x + 2y - 3z - 6 = 0 \end{cases} \end{aligned}$$

$$\text{Sätt:} \quad P(\boldsymbol{x}, r) = f(\boldsymbol{x}) + \frac{1}{r}\left(g_1(\boldsymbol{x})^2 + g_2(\boldsymbol{x})^2\right)$$

Optimering med bivillkor

Straffmetoder för icke-linjära olikhetsvillkor

För $f : \mathbb{R}^n \to \mathbb{R}$ och $g : \mathbb{R}^n \to \mathbb{R}^m$ och optimeringsproblemet

$$(P_I) \quad \begin{array}{ll} \min & f(x) \\ \text{s.t.} & g(x) \leq 0 \end{array}$$

inför vi en yttre straffunktion

$$P(x, r) = f(x) + \frac{1}{r} g^+(x)^T g^+(x) \quad \text{där} \quad g_i^+(x) = \begin{cases} g_i(x) & \text{om} \quad g_i(x) > 0 \\ 0 & \text{annars} \end{cases}$$

- Bestraffning sker om och endast om punkten x ligger utanför det tillåtna området (d.v.s., när $g_i(x) > 0$ för något i).
- Metoden kallas yttre straffmetod eftersom optimallösningen x^* närmas utifrån det tillåtna området ($x^*(r) \to x^*$ då $r \to 0^+$) genom att lösa $\min P(x, r)$ för successivt mindre r.

Optimering med bivillkor

Straffmetoder för icke-linjära olikhetsvillkor

- Om det inte är möjligt att beräkna f eller g utanför det tillåtna området kan en inre straffmetod (eller barriärmetod) användas. Ex:
  1. Invers barriärfunktion: $P(x,r) = f(x) - r\sum_{i=1}^{m} \frac{1}{g_i(x)}$
  2. Logaritmisk barriärfunktion: $P(x,r) = f(x) - r\sum_{i=1}^{m} \ln(-g_i(x))$

  Båda har en "barriär" vid randen av det tillåtna området
- När x närmar sig gränsen till det tillåtna området ($g_i(x) \to 0^+$):
  - Invers barriär: $-r \cdot (1/g_i(x)) \to +\infty$
  - Logaritmisk barriär: $-r \cdot \ln(-g_i(x)) \to +\infty$

  Detta skapar en oändligt hög kostnad (barriär) vid gränsen vilket effektivt hindrar optimeringsalgoritmen från att nå eller korsa det icke-tillåtna området.
- $x^*(r) \to x^*$ inifrån när straffparametern $r \to 0^+$.
- Inre barriärmetoder håller alltid sökningen efter optimum strikt innanför det tillåtna området.

Optimering med bivillkor

Straffmetoder för icke-linjära olikhetsvillkor

Exempel 11

Bestäm

$$\begin{array}{ll} \min & f(x) = x^3 + x \\ \text{s.t.} & x \geq 2 \end{array}$$

med inre straffmetod och invers barriärfunktion.
Vi har olikhetsvillkoret $g(x) = 2 - x \leq 0$.

$$\text{Straffunktion:} \quad P(x, r) = f(x) - r\frac{1}{g(x)} = x^3 + x - r\frac{1}{2 - x}.$$

$$\text{Stationära punkter:} \quad \begin{aligned} \frac{dP}{dx} &= 3x^2 + 1 - r\frac{1}{(2 - x)^2} = 0 \Leftrightarrow r = (3x^2 + 1)(2 - x)^2 \\ \therefore \quad & r \to 0^+ \Leftrightarrow x \to 2^+ \end{aligned}$$

Optimering med bivillkor

Straffmetoder för icke-linjära olikhetsvillkor

Exempel 12

Samma problem igen: Bestäm

$$\begin{array}{ll} \min & f(x) = x^3 + x \\ \text{s.t.} & x \geq 2 \end{array}$$

med inre straffmetod och logaritmisk barriärfunktion.
Vi har fortfarande olikhetsvillkoret $g(x) = 2 - x \leq 0$.

Straffunktion: $P(x, r) = f(x) - r\ln(-g(x)) = x^3 + x - r\ln(-(2 - x)).$

Stationära punkter: $$\begin{aligned} \frac{dP}{dx} &= 3x^2 + 1 - r\frac{1}{x-2} = 0 \Leftrightarrow r = (3x^2+1)(x-2) \\ \therefore &\quad r \to 0^+ \Leftrightarrow x \to 2^+ \end{aligned}$$